

## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/79746>

Please be advised that this information was generated on 2017-12-06 and may be subject to change.

Research article

Open Access

# A simple method for co-segregation analysis to evaluate the pathogenicity of unclassified variants; *BRCA1* and *BRCA2* as an example

Leila Mohammadi<sup>†1</sup>, Maaïke P Vreeswijk<sup>†2</sup>, Rogier Oldenburg<sup>4</sup>, Ans van den Ouweland<sup>4</sup>, Jan C Oosterwijk<sup>5</sup>, Annemarie H van der Hout<sup>5</sup>, Nicoline Hoogerbrugge<sup>6</sup>, Marjolijn Ligtenberg<sup>6</sup>, Margreet G Ausems<sup>7</sup>, Rob B van der Luijt<sup>7</sup>, Charlotte J Dommering<sup>8</sup>, Johan J Gille<sup>8</sup>, Senno Verhoef<sup>9</sup>, Frans B Hogervorst<sup>9</sup>, Theo A van Os<sup>10</sup>, Encarna Gómez García<sup>11,12</sup>, Marinus J Blok<sup>12</sup>, Juul T Wijnen<sup>2</sup>, Quinta Helmer<sup>1</sup>, Peter Devilee<sup>2,3</sup>, Christi J van Asperen<sup>\*2</sup> and Hans C van Houwelingen<sup>1</sup>

Address: <sup>1</sup>Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands, <sup>2</sup>Center for Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands, <sup>3</sup>Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands, <sup>4</sup>Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands, <sup>5</sup>Department of Genetics, University Medical Center, Groningen University, Groningen, The Netherlands, <sup>6</sup>Department of Human Genetics, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands, <sup>7</sup>Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands, <sup>8</sup>Department of Clinical Genetics, VU University Medical Center, Amsterdam, The Netherlands, <sup>9</sup>Family Cancer Clinic, Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands, <sup>10</sup>Department of Clinical Genetics, Academic Medical Center, Amsterdam, The Netherlands, <sup>11</sup>Department of Clinical Genetics, University Medical Center, Maastricht, The Netherlands and <sup>12</sup>Department of Genetics and Cell Biology, University Medical Center, Maastricht, The Netherlands

Email: Leila Mohammadi - l.mohammadikhankahdani@erasmusmc.nl; Maaïke P Vreeswijk - vreeswijk@lumc.nl; Rogier Oldenburg - r.oldenburg@erasmusmc.nl; Ans van den Ouweland - a.vandenouweland@erasmusmc.nl; Jan C Oosterwijk - j.c.oosterwijk@medgen.umcg.nl; Annemarie H van der Hout - A.H.van.der.Hout@medgen.azg.nl; Nicoline Hoogerbrugge - n.hoogerbrugge@antrg.umcn.nl; Marjolijn Ligtenberg - m.ligtenberg@antrg.umcn.nl; Margreet G Ausems - M.G.E.M.Ausems@umcutrecht.nl; Rob B van der Luijt - r.b.vanderluijt@umcutrecht.nl; Charlotte J Dommering - cj.dommering@vumc.nl; Johan J Gille - jip.gille@vumc.nl; Senno Verhoef - s.verhoef@nki.nl; Frans B Hogervorst - f.hogervorst@nki.nl; Theo A van Os - t.a.vanos@amc.uva.nl; Encarna Gómez García - Encarna.Gomezgarcia@gen.unimaas.nl; Marinus J Blok - r.blok@gen.unimaas.nl; Juul T Wijnen - J.Wijnen@lumc.nl; Quinta Helmer - q.helmer@lumc.nl; Peter Devilee - p.devilee@lumc.nl; Christi J van Asperen\* - asperen@lumc.nl; Hans C van Houwelingen - J.C.van\_Houwelingen@lumc.nl

\* Corresponding author †Equal contributors

Published: 29 June 2009

Received: 24 November 2008

BMC Cancer 2009, 9:211 doi:10.1186/1471-2407-9-211

Accepted: 29 June 2009

This article is available from: <http://www.biomedcentral.com/1471-2407/9/211>

© 2009 Mohammadi et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

**Background:** Assessment of the clinical significance of unclassified variants (UVs) identified in *BRCA1* and *BRCA2* is very important for genetic counselling. The analysis of co-segregation of the variant with the disease in families is a powerful tool for the classification of these variants. Statistical methods have been described in literature but these methods are not always easy to apply in a diagnostic setting.

**Methods:** We have developed an easy to use method which calculates the likelihood ratio (LR) of an UV being deleterious, with penetrance as a function of age of onset, thereby avoiding the use of

liability classes. The application of this algorithm is publicly available <http://www.msbi.nl/cosegregation>. It can easily be used in a diagnostic setting since it requires only information on gender, genotype, present age and/or age of onset for breast and/or ovarian cancer.

**Results:** We have used the algorithm to calculate the likelihood ratio in favour of causality for 3 UVs in *BRCA1* (p.M18T, p.S1655F and p.R1699Q) and 5 in *BRCA2* (p.E462G, p.Y2660D, p.R2784Q, p.R3052W and p.R3052Q). Likelihood ratios varied from 0.097 (*BRCA2*, p.E462G) to 230.69 (*BRCA2*, p.Y2660D). Typing distantly related individuals with extreme phenotypes (i.e. very early onset cancer or old healthy individuals) are most informative and give the strongest likelihood ratios for or against causality.

**Conclusion:** Although co-segregation analysis on itself is in most cases insufficient to prove pathogenicity of an UV, this method simplifies the use of co-segregation as one of the key features in a multifactorial approach considerably.

## Background

High throughput technologies and more sensitive mutation detection systems in the DNA diagnostic laboratories have led to an increasing number of sequence variants in the major cancer-predisposing genes for which the clinical significance is unknown. In approximately 15% of the DNA mutation scannings of *BRCA1* or *BRCA2* in breast cancer families, the test result is difficult to interpret because an unclassified variant is found (UV, or 'variant of uncertain clinical significance' (VUCS) or 'variant of uncertain significance' (VUS) [1]. The risk for breast and ovarian cancer might be as high as for classical pathogenic mutations, but they might also be negligible. The decision for or against prophylactic surgery must in these cases be based entirely on the family cancer history, which never yields risks comparable to pathogenic mutations. In addition, predictive testing and preventive surgery are not offered to their healthy relatives, although these relatives may experience anxiety for carrying the UV [2].

A variety of approaches have been used to assess the clinical relevance of these UVs [3-9]. These include 1) *in silico* predictions based on evolutionary conservation, position and nature of the amino acid change, 2) functional analysis of the variant using *in vitro* assays directed at specific functional domains of the protein, 3) population genetics analyzing frequency of the variant in cases and controls and co-occurrence of the UV with a known deleterious mutation in one or more tested individuals and 4) clinical validation using family history, co-segregation of the variant with the disease in pedigrees as well as relevant features of *BRCA1*- and *BRCA2*-associated tumours.

A functional assay would ideally be capable of distinguishing pathogenic from neutral variants. However, although these assays are available, their results are very difficult to interpret clinically and can only be used for a subset of unclassified variants located in specific domains of the genes. This is because *BRCA1* and *BRCA2* are multifunctional proteins and these assays will generally only

interrogate one of those functions. Since none of the approaches mentioned above have by itself been able to provide compelling evidence for or against pathogenicity of UVs, multifactorial likelihood models integrating these features have been developed [3,4]. As shown by Goldgar *et al.*, the significance of the analysis was highly dependent on co-occurrence and co-segregation data. Recently, the genetic evidence for or against disease causality was reported for a large number of variants with the use of a large database of tested individuals collected analyzing co-occurrence, personal and family history of cancer as well as co-segregation with disease, the latter only in a subset of probands [5]. Although co-occurrence has shown to be a powerful tool to obtain high odds in favour of neutrality, this method requires large databases of tested individuals, such as collected by Myriad Genetics laboratories, and these data are not publicly available. In addition, family-history analysis has demonstrated to be very useful although it might be susceptible to bias due to population-specific ascertainment. Co-segregation (occurrence of UVs in affected individuals) is not affected by selection bias because it considers the distribution of the genotypes given the phenotypes and the genotype of the proband (the first person that was genotyped in the family and carried a UV). Co-segregation analysis is therefore regarded as a robust approach since it directly relates to the disease risk whereas the absence of co-segregation provides strong evidence against pathogenicity.

In order to classify UVs via co-segregation, a simple Bayesian method to assess causality of rare sequence variants was provided by Petersen *et al.* [10]. In contrast to Petersen, Thompson *et al.* [11] have developed a more general method based on the full pedigree likelihood. All available genotype information from the family is used, including any unaffected individuals who have been tested. The two methods either use a defined penetrance in carriers *versus* non-carriers and ignore the age of onset [11], or specify liability classes which define the age range of family members in intervals for which the penetrance is

supposed to be constant [10]. Using the full age of onset information however will result in more reliable estimates of the likelihood ratio (LR).

We have developed an algorithm to calculate the likelihood ratio of a UV being deleterious based on co-segregation analysis, using the precise age of onset information. The method can easily be used in a diagnostic setting since it requires only information on gender, genotype, present age and/or age of onset for breast and/or ovarian cancer.

## Methods

In collaboration with the DNA diagnostic laboratories of the Clinical Genetic Centres in the Netherlands, 8 UVs have been selected for this study (Table 1, detailed description can be found in Additional file 1; Table S1). The Human Genetic Variation Society (HGVS) approved guidelines <http://www.hgvs.org/mutnomen> have been used for *BRCA1* and *BRCA2* nomenclature [12]. To facilitate published data comparison, also the traditional nomenclature is listed (Breast Cancer Information Core, <http://research.nhgri.nih.gov/bic/>). GenBank accession no. [NM\\_007294.2/NP\\_009226.1](#) and [NM\\_000059.3/NP\\_000050.1](#) have been used for *BRCA1* and *BRCA2* mRNA and protein numbering respectively.

Each laboratory selected 5 UVs in either *BRCA1* or *BRCA2* which were of particular interest for them (e.g. multiple families with same UV, large families in which the UV was segregating). A long-list was made which contained all the

selected UVs (14 *BRCA1*, 17 *BRCA2*). From this list, UVs were selected that met a least one of the following criteria: 1) Grantham score above 100, 2) The UV is located in a functional domain [13], 3) The amino acid has a high degree of evolutionary conservation (i.e. also in non-mammalian species). In addition, priority was given to those UVs found in more than one family and/or genetic centre. UVs co-occurring with a pathogenic mutation in the same gene (data from Netherlands) in the proband were excluded. 15 Dutch families, in which one of the 8 selected UVs had been found and for which at least two family members were genotyped were included in our study. Some of the UV's were detected in more than one family. Families were tested for *BRCA1/2* after genetic counselling of clinically presumed hereditary breast and/or ovarian cancer, when the mutation detection rate is above 10% [14] (e.g., two breast cancer cases with one case under age 50, more than 3 first-degree relatives in two successive generations with breast cancer under age 60, breast and/or ovarian cancer families), or if breast cancer was diagnosed at a relatively young age (i.e., one patient younger than 35 years old). The proband is defined as the first person in the family who is genotyped and tested positive for one of the selected UVs. The pathogenic control group consisted of families in which a pathogenic mutation was detected in *BRCA1* or *BRCA2* (Table 2). The neutral control group consisted of families in which genetic variants were found that are considered neutral variants with respect to breast and/or ovarian cancer risk (Table 3) [3,15].

**Table 1: *BRCA1* and *BRCA2* missense variants analysed and likelihood ratio in favor of causality**

Gene	HGVS <sup>1</sup>	traditional <sup>2</sup>	protein	# fam <sup>3</sup>	LR per family <sup>4</sup>	LR-variant <sup>5</sup>
<i>BRCA1</i>	c.53T>C	172T>C	p.M18T	2	5.7004(2 <sup>+</sup> ) 1.3995(4 <sup>+</sup> ; 6 <sub>+</sub> ; 3 <sub>+</sub> )	7.9777
<i>BRCA1</i>	c.4964C>T	5083C>T	p.S1655F	1	6.7425(3 <sup>+</sup> ; 2 <sub>+</sub> )	6.7425
<i>BRCA1</i>	c.5096G>A	5215G>A	p.R1699Q	1	1.4280(2 <sup>+</sup> )	1.4280
<i>BRCA2</i>	c.1385A>G	1613A>G	p.E462G	4	0.7738(1 <sup>+</sup> ; 1 <sub>+</sub> ) 0.6757(1 <sup>+</sup> ; 1 <sub>+</sub> ) 1.2459(1 <sub>+</sub> ; 1 <sub>+</sub> ) 0.1481(1 <sup>+</sup> ; 1 <sub>+</sub> )	0.0965
<i>BRCA2</i>	c.7978T>G	8206T>G	p.Y2660D	3	1.8579(2 <sup>+</sup> ), 11.1565(2 <sup>+</sup> ) 11.1297(2 <sup>+</sup> ; 3 <sub>+</sub> )	230.6927
<i>BRCA2</i>	c.8351C>T	8579G>A	p.R2784Q	1	1.8101(1 <sup>+</sup> ; 1 <sub>+</sub> )	1.8101
<i>BRCA2</i>	c.9154C>T	9382C>T	p.R3052W	2	3.0651(2 <sup>+</sup> ), 3.9790(3 <sup>+</sup> )	12.1960
<i>BRCA2</i>	c.9155G>A	9383G>A	p.R3052Q	1	0.2208(1 <sup>+</sup> ; 1 <sub>+</sub> )	0.2208

<sup>1</sup>The Human Genetic Variation Society (HGVS) approved guidelines <http://www.hgvs.org/mutnomen> have been used for *BRCA1* and *BRCA2* nomenclature [12].

<sup>2</sup>To facilitate published data comparison, also the traditional nomenclature is listed (Breast Cancer Information Core, <http://research.nhgri.nih.gov/bic/>). GenBank accession no. [NM\\_007294.2/NP\\_009226.1](#) and [NM\\_000059.3/NP\\_000050.1](#) have been used for *BRCA1* and *BRCA2* mRNA and protein numbering respectively.

<sup>3</sup>Number of families with at least two genotyped family members.

<sup>4</sup>Likelihood ratio per family. Between brackets the number of affected individuals carrying the UV (n<sup>+</sup>) or unaffected individuals with the UV (n<sub>+</sub>), affected individuals without the UV (n<sup>-</sup>) or unaffected individuals without the UV (n<sub>-</sub>).

<sup>5</sup>The overall likelihood ratio is derived by the product of the LR over the independent families.

**Table 2: Likelihood ratio in favor of causality for selected pathogenic mutations**

Gene	HGVS <sup>1</sup>	traditional <sup>2</sup>	protein	# fam <sup>3</sup>	LR per family <sup>4</sup>	LR-variant <sup>5</sup>
BRCA1	c.81-6T>A	IVS2-6T>A	p.Cys27fsX1	1	6.4287(5 <sup>+</sup> ; 3 <sub>+</sub> ; 1 <sub>+</sub> ; 3 <sub>+</sub> )	6.4287
BRCA1	c.213-12A>G	IVS5-12A>G	p.Arg71fsX20	1	10.2701(1 <sup>+</sup> ; 3 <sub>+</sub> ; 6 <sub>+</sub> )	10.2701
BRCA1	c.1292dup	1411insT	p.Leu431PhefsX5	2	1.7312(2 <sup>+</sup> ; 1 <sub>+</sub> ; 1 <sub>+</sub> ) 1.9475(1 <sup>+</sup> ; 2 <sub>+</sub> ; 1 <sub>+</sub> )	3.3715
BRCA1	c.2193_2197del	2312delAGAAG	p.Glu733ThrfsX5	1	4.7089(2 <sup>+</sup> ; 1 <sub>+</sub> ; 4 <sub>+</sub> )	4.7089
BRCA1	c.5095C>T	5214C>T	p.Arg1699Trp	3	22.8602(3 <sup>+</sup> ; 6 <sub>+</sub> ; 2 <sub>+</sub> ; 4 <sub>+</sub> ) 2.2720(1 <sup>+</sup> ; 1 <sub>+</sub> ; 1 <sub>+</sub> ) 7.7431(2 <sup>+</sup> ; 1 <sub>+</sub> )	402.164
BRCA2	c.2806_2809del	3034delAAAC	p.Ala938ProfsX21	1	1.9901(2 <sup>+</sup> ; 1 <sub>+</sub> ; 1 <sub>+</sub> )	1.9901
BRCA2	c.3269del	3497delT	p.Met1090SerfsX14	1 <sup>6</sup>	87.46(5 <sup>+</sup> )	87.46
BRCA2	c.3599_3560del	3827delTGT	p.Cys1200X	1	1.1876(1 <sup>+</sup> ; 3 <sub>+</sub> ; 1 <sub>+</sub> )	1.1876
BRCA2	c.8067T>A	8295T>A	p.Cys2689X	1	12.2424(3 <sup>+</sup> ; 1 <sub>+</sub> ; 4 <sub>+</sub> )	12.2424
BRCA2	c.8773C>T	9001C>T	p.Gln2925X	1	1.8827(1 <sup>+</sup> ; 3 <sub>+</sub> )	1.8827

<sup>1-5</sup> For clarification see Table 1.<sup>6</sup> Pedigree is shown in Figure 2.**Likelihood ratio using co-segregation data**

In this subsection, we derive the relevant likelihood ratio to evaluate the evidence for the causality associated with the UV referred to breast or ovarian cancer. The first assumption is that the proband carries a UV in *BRCA1* or *BRCA2*.

The notation G stands for the genotype of the UV in the family. Hence, G = 1 if the UV is present and G = 0 if not.

The genotype is observed in at least one other family member, but certainly not in all. Phenotypic information is available in most family members and denoted by P. Pedigrees are pruned to leave out non-informative branches (i.e. no cancer and no genotype information).

We use the notations "d", "n", "o", "u", "f", "p" for "deleterious", "neutral", "observed", "unobserved", "family" and "proband" respectively. The likelihood ratio for pathogenicity based on segregation within families is then equal to

$$LR = \frac{P_d(G_0|P_f, G_p=1)}{P_n(G_0|P_f, G_p=1)}$$

By definition of "neutral"

$$P_n(G_0 | P_f, G_p = 1) = P(G_0 | G_p = 1).$$

One has

$$P_d(G_0 | P_f, G_p = 1) = \frac{P_d(P_f|G_0, G_p=1)P(G_0, G_p=1)}{\sum_{G'_0} P_d(P_f|G'_0, G_p=1)P(G'_0, G_p=1)}$$

Hence,

$$LR = \frac{P_d(P_f|G_0, G_p=1)}{\sum_{G'_0} P_d(P_f|G'_0, G_p=1)P(G'_0, G_p=1)}.$$

We have to take into account all family members with unobserved (missing) genotypic information. If we do so, we obtain

$$\begin{aligned} LR &= \frac{\sum_{G_u} P_d(P_f|G_u, G_0, G_p=1)P(G_u, G_0|G_p=1)}{\sum_{G_f} P_d(P_f|G_f)P(G_f|G_p=1)} \\ &= \frac{\sum_{G_u} P_d(P_f|G_f)P(G_f|G_p=1)}{\sum_{G_f} P_d(P_f|G_f)P(G_f|G_p=1)} \end{aligned}$$

**Table 3: Likelihood ratio in favor of causality for selected neutral variants**

Gene	HGVS <sup>1</sup>	traditional <sup>2</sup>	protein	# fam <sup>3</sup>	LR per family <sup>4</sup>	LR-variant <sup>5</sup>
BRCA1	c.135-15_135-12del <sup>6</sup>	IVS4-15delCTTT	p.=	1 <sup>9</sup>	3.8592 (2 <sup>+</sup> )	3.8592
BRCA1	c.2613G>A <sup>7</sup>	2732G>A	p.Pro871Pro	1	1.8533(2 <sup>+</sup> )	1.8533
BRCA1	c.5152+20T>A <sup>7</sup>	IVS18+20T>A	p.=	1	3.6711(1 <sup>+</sup> ; 1 <sub>+</sub> )	3.6711
BRCA2	c.125A>G <sup>8</sup>	353A>G	p.Y42C	2	0.9728(2 <sub>+</sub> ) 0.7612(1 <sup>+</sup> ; 4 <sub>+</sub> )	0.7405

<sup>1-5</sup> For clarification see Table 1.These variants are regarded to be neutral towards cancer risk based on <sup>6</sup>Vreeswijk et al. [15], <sup>7</sup>Splice site prediction analysis and <sup>8</sup>Goldgar et al. [3].<sup>9</sup> Pedigree is shown in Additional file 1; Figure S4.

So, the building blocks are  $P(G_f|G_p = 1)$  and  $P_d(P_f|G_f)$ .

We can now make some simplifying assumptions

a. the mutation is very rare. That helps enormously in enumerating all possible genotypes in the family and computing  $P(G_f|G_p = 1)$

b. if the mutation is deleterious and known to be present in the family, the genotype for that mutation is the dominant factor in computing  $P_d(P_f|G_f)$ , that is to say that the probability of the phenotype for an individual only depends on her/his genotype and not on the phenotypes of the other family members. This implies that  $P_d(P_f | G_f f) = \prod_i P_d(P_i | G_i)$ , where  $i$  changes over all family members.

If the mutation is deleterious, we assume that the penetrance of the mutation (defined by the probability of acquiring the disease before age  $t$ ) is given by the known penetrance for *BRCA1* or *BRCA2*, respectively. We apply the penetrances as in Jonker *et al.* [16], the smoothed versions of the known penetrances based on normal distribution functions. We denote the density and cumulative functions of normal distribution by  $\phi$  and  $\Phi$  respectively. The general form is given by  $F(t|r, \mu, \sigma) = r \Phi((t - \mu)/\sigma)$ . Values of  $r$  (life time risk),  $\mu$  (age of diagnosis) and  $\sigma$  (standard deviation) for the breast and ovarian cancer are given in Additional file 1; Table S2. The penetrance functions are plotted in Additional file 1; Figure S1. The phenotypic likelihood is given by  $1 - F(t|r, \mu, \sigma)$  if an individual is disease free at age  $t$ , and by the derivative  $f(t|r, \mu, \sigma) = r\phi(t - \mu)/\sigma$  if the individual has cancer at age  $t$ . Breast and ovarian cancer are assumed to be statistically independent given the genotype of the deleterious mutation.

The likelihood for two sided breast cancer cases at onset ages  $t_1$  and  $t_2 (\geq t_1)$  is given as the product of the derivatives of  $1 - \sqrt{1 - F(t_1 | r, \mu, \sigma)}$  and  $1 - \sqrt{1 - F(t_2 | r, \mu, \sigma)}$  [17].

For a specific variant, if there are multiple families carrying the variant, the overall likelihood is derived by the product of the likelihood ratios over the independent families, assuming that the specific variant is not in linkage disequilibrium with an unknown deleterious mutation.

#### An algorithm to enumerate possible genotypes

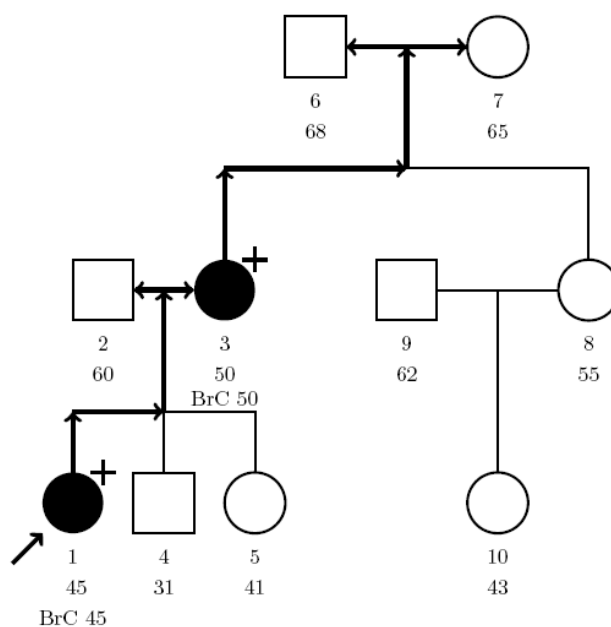
To analyze co-segregation and to compute  $P(G_f|G_p = 1)$  under the rare genotype assumption, one has to explore

all the different possibilities of the genotypes of the family members. The genotype takes values 1 for the carriers of the variant and 0 for non-carriers. In our data set the proband is by definition carrier of a variant (genotype 1).

The algorithm that is at the basis of our computational algorithm is demonstrated using the pedigree of Figure 1.

The members of the pedigree are numbered  $1, \dots, n$ . In our example, the proband is given number 1. A genotypic configuration is described by a vector  $G = (G_1, \dots, G_n)$  of 0's and 1's,  $G_i = 1$  if individual  $i$  carries the UV and  $G_i = 0$  if it is a non-carrier. Under the rare mutation assumption and the selection on the proband, all the possibilities have to meet the following

1.  $G_p = 1$ .
2. If two parents have  $G = 0$ , all children have  $G_i = 0$ .
3. If for an individual  $G_i = 1$ , one (and only one) of his/her parents must have the mutation, i.e.  $G = 1$ .
4. Exactly one of the founders has  $G = 1$ .



**Figure 1**

**Illustration of our model.** Hypothetical pedigree for the illustration of our model. Individuals are numbered 1–10 for identification and underneath the age at last contact is listed. Open circle: female, open square: male, closed circle: affected with breast (or ovarian) cancer at age  $x$  (BrCx). Unless specified by + (carrier of UV), - (no carrier), individuals are not genotyped. Proband is indicated by arrow.

Conditions 3 and 4 are equivalent formulations of the rare mutation assumption.

In order to construct an enumeration of all possible genotypic configurations, we proceed as follows

#### Step 1

##### Finding possible founders

Going upwards from the proband we can find all founders that can be reached from the proband. Given a founder, all family members in the path from the founder to proband must have the mutation as well. All the pedigree members that do not descend from the founder with the mutation cannot have the mutation.

In the example (Figure 1), person 1 is the proband carrying the UV. There are three possible founders: person 2, 6 or 7. We will discuss these three possibilities below.

1) If person 2 carries the mutation, then numbers 3, 6, 7, 8 and 10 can not have the mutation. Person 4 and 5 may have the mutation. We first indicate the genotype of numbers 4 and 5 by  $x$  and then obtain all possibilities for  $x$ 's. So all the possibilities in case that number 2 carries the mutation are contained in the following vector where the upper line represents the number of the person in the pedigree and the lower line represents the genotype information.

$$M_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 1 & 1 & 0 & x & x & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

2) If number 7 has the mutation, then numbers 2 and 6 cannot have the mutation. Number 3 will carry the mutation and numbers 4, 5, 8 and 10 may have the mutation. Again we first indicate the genotype of numbers 4, 5, 8 and 10 by  $x$  and then obtain all possibilities for  $x$ 's. So all the possibilities for the case that number 7 has the mutation are contained in the following vector

$$M_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 1 & 0 & 1 & x & x & 0 & 1 & x & 0 & x \end{pmatrix}$$

3) If number 6 has the mutation, then we repeat the situation 2) above with numbers 6 and 7 replaced by each other.

Hence, all the possible configurations are contained in the vectors  $M_1$  and  $M_2$ .

The situations 1), 2) and 3) are also shown in Additional file 1; Figure S2A. This figure might be useful to understand how the algorithm is built.

The  $x$ 's have to be filled to obtain the enumeration. To do this in an efficient way we have to determine the "generation" of the  $x$ 's.

#### Step 2

##### Setting the generations.

We use the following iterative algorithm to define the "generation"  $H$  for the

"open" genotypes:

- if both parents have known genotype, then  $H = 2$ ,
- if one parent has "open" genotype (two is impossible), then  $H_i = H_{parent} + 1$ .

We fill in the generations at the "open" genotypes. We obtain

$$M_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 1 & 1 & 0 & 2 & 2 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$M_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 1 & 0 & 1 & 2 & 2 & 0 & 1 & 2 & 0 & 3 \end{pmatrix}$$

#### Step 3

##### Filling the "hole"

The general principle is that an open genotype must be "0" if both parents have  $G_i = 0$  and can be 0 or 1 if one of the parents has  $G_i = 1$ . That leads to the filling and splitting algorithm shown in Additional file 1; Table S3.

Additional file 1; Figure S2B shows the situation 1) above. As mentioned above, all possibilities in this situation are contained in vector  $M_1$ .

We start with this vector and we need to find all different possibilities for the 2's in this vector. We start with the member number 4.

Because one of the parents has the mutation, number 4 can have  $G = 0$  and  $G = 1$ . For each of these possibilities, (with the same reason) number 5 also has two possibilities  $G = 0$  and  $G = 1$ . In fact when a parent has  $G = 1$ , there is a split, otherwise there is no split.

The situation 2) above is shown in Additional file 1; Figure S3. As mentioned above, all possibilities in this situation are contained in vector  $M_2$ .

We start with this vector and we need to find all different possibilities for the 2's and 3's in this vector. We start with the member number 4. As our explanation for Additional

file 1; Figure S2B, we have a split for number 4. For each of the possibilities, we have another split for number 5 and for each of them; we have another split for number 8. Then all the 2's are filled and we arrive at the 3 for number 10. Here we have a split just for the possibilities that number 8 has  $G = 1$  otherwise there is no splits and number 10 has  $G = 0$ .

Using our algorithm, the two matrixes  $M_1$  and  $M_2$  become

$$M_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$M_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

To compute  $P(P_f | G_p = 1)$  one needs to derive the probability of each possibility.

Note that

$$P(G_{father} = 0, G_{mother} = 1 | G_{child} = 1) =$$

$$P(G_{father} = 1, G_{mother} = 0 | G_{child} = 1) = 1/2$$

and

$$P(G_{child} = 1 | G_{father} = 1, G_{mother} = 0) =$$

$$P(G_{child} = 1 | G_{father} = 0, G_{mother} = 1) = 1/2$$

So, each split halves the probability. This is shown in Additional file 1; Figure S2.

The probability of each possibility is the product of the  $1/2$ 's. In fact a split occurs when a member of the family has a parent with  $G = 1$ . Hence, the probability of a possibility

is  $1/2$  to the power of the number of family members who have one parent with  $G = 1$  in the possibility.

For the two matrixes,  $M_1$  and  $M_2$ , we obtain the probabilities below for the respective rows.

$$[2^{-3}, 2^{-3}, 2^{-3}, 2^{-3}],$$

$$[2^{-6}, 2^{-6}, 2^{-5}, 2^{-6}, 2^{-6}, 2^{-5}, 2^{-6}, 2^{-6}, 2^{-5}, 2^{-6}, 2^{-6}, 2^{-5}].$$

Assuming that the proband in Figure 1 carries a *BRCA1* mutation, the likelihood ratio in favour of causality is 1.66. The likelihood ratio becomes 1.76 in case of a *BRCA2* mutation, due to differences in penetrance between both genes.

The algorithm can be run on the website at <http://www.msbi.nl/cosegregation>. On the website, users can upload a file containing the pedigree information and the application will calculate the likelihood ratio instantly. A detailed instruction for the use of the method is available as Additional file 2 as well as on the website.

## Results

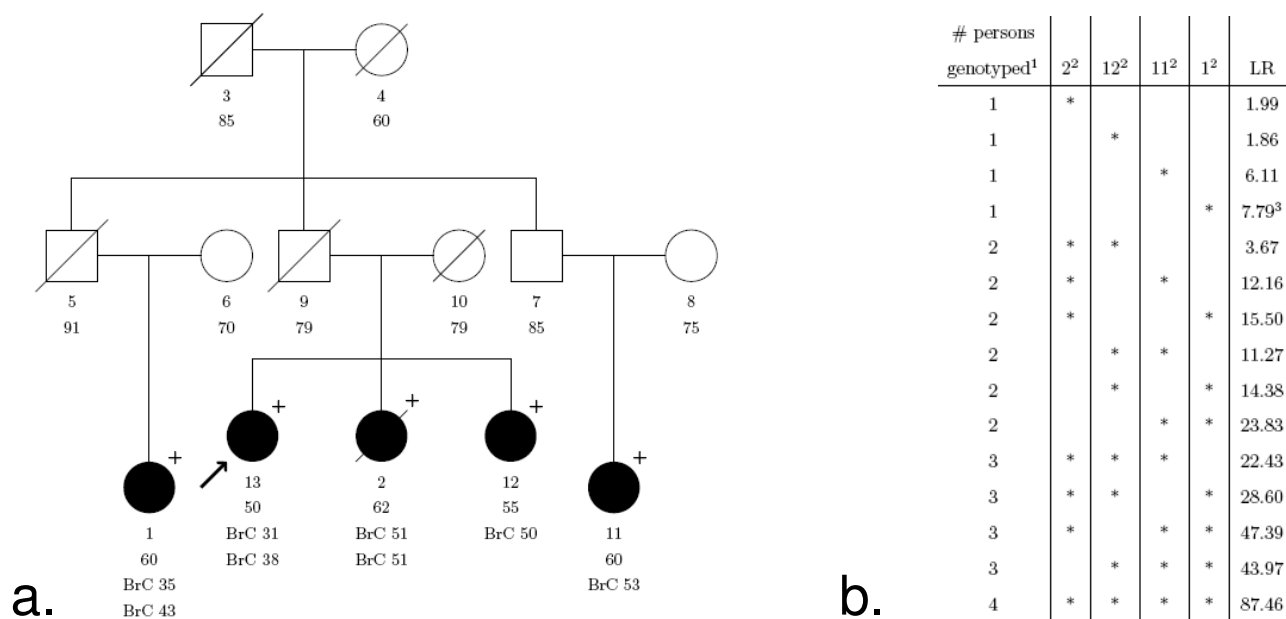
We have calculated the likelihood ratios for 8 UVs in *BRCA1* and *BRCA2* using the method described in the Methods section. In Table 1 the likelihood ratios are given for the separate UV-families, as well as the combined odds for an UV for which multiple families were available. Likelihood ratios for families with pathogenic mutations and neutral variants are summarized in Table 2 and 3 respectively.

The multifactorial likelihood model described by Goldgar *et al.* [3], considered UVs with overall odds of causality  $>1000:1$  ( $LR > 1000$ ) as pathogenic and those with odds of causality  $<1:100$  ( $LR < 0.01$ ) as neutral.

Except for the *BRCA2* p.Y2660D variant, with a LR in favour of causality of more than 230, none of the UV's has a very outspoken LR in favour of causality ( $LR > 1$ ) or neutrality ( $LR < 1$ ). Moreover, there is quite some heterogeneity between families. Similarly, we see that the evidence for the pathogenic and neutral controls is not very strong either. In this section we describe some further data-analysis to get a better insight into the usefulness of co-segregation in the classification of UV's. Moreover, we want to explore what the best genotyping strategy should be: Should we genotype close or distant relatives of the proband and patients or healthy individuals?

To explore the latter issue we study a pedigree with a pathogenic *BRCA2* mutation (c.3269del with a LR of 87.46) in more detail (Figure 2). We calculated what the LR would be if we had genotyped only a subset of the family mem-





**Figure 2**  
**Likelihood ratio in favour of causality for different genotyping patterns.** a. Pedigree with *BRCA2* mutation c.3269del. Individuals are numbered 1–13 for identification and underneath the age at last contact is listed. Open circle: female, open square: male, /: deceased, closed circle: affected with breast (or ovarian) cancer at age × (BrCx). In case of bilateral breast cancer this is listed below the first occurrence. Unless specified by + (carrier of UV), - (no carrier), individuals are not genotyped. Proband is indicated by arrow. b. Likelihood ratios in favour of causality if only a subset of the family members are genotyped and in different combinations. <sup>1</sup>The number of family members that is genotyped positive for the mutation, in addition to the proband, individual 13, who carries the *BRCA2* mutation c.3269del mutation. <sup>2</sup>Family member number in pedigree. <sup>3</sup>Likelihood ratio in favour of causality when individual 13 and 1 carry the mutation.

bers and in different combinations. Typing for example one sister of the proband with breast cancer at age 50 (individual 12) will give a LR of 1.86 when she carries the mutation, whereas a positive genotype (i.e. carrier of the UV) of a third grade related cousin with breast cancer at age 53 (individual 11) results in a LR of 6.11. Two affected sisters with the UV or two affected cousins will generate a LR of 3.67 and 23.83 respectively. From this analysis it is clear that typing distantly related individuals with extreme phenotypes (i.e. very early onset cancer or old healthy individuals) is most informative and gives the strongest likelihood ratios for or against causality. Small pedigrees with only one or two affected individuals do not carry very much information that helps to decide on the clinical significance of the UV. Generally speaking it is helpful to genotype as many individuals as possible. If it is only possible to type a limited number of individuals, the advice is to type individuals with "extreme phenotypes" in these breast cancer families. Distant affected young relatives which are carriers and nearby unaffected old non carriers (>60 years of age) are a very strong indication for causality of the variant whereas variants found in distant unaffected old family members and not in nearby affected young members will very likely be neutral to cancer risk.

To explore the sensitivity of the method we have to look into one of the pedigrees with a neutral variant (*BRCA1* c.135-15\_135-12del; Additional file 1; Figure S4). We distributed the rare variant randomly from the proband onwards using the algorithm described in the Methods section and looked what distribution of the LR we would get depending on who we genotyped and what the outcome would be. If we would type everybody, we would obtain a distribution of the Likelihood Ratio for which the probability of getting a LR>1 is almost 20% and the probability of getting LR>3 almost 8%. The probability of obtaining a LR<0.1 in favour of neutrality, is also almost 8% (Additional file 1; Figure S5).

One might wonder if it would not have been wiser to genotype another pedigree member. However, a more detailed analysis of the simulation data (not shown) reveals that the LR of 2.97 that we get if we only type individual 4 is no indication that the wrong individual was genotyped. Generally, individuals that show strong evidence when their genotype is positive also show strong evidence when the genotype is negative.

## Discussion

To assess the clinical relevance of unclassified variants, a number of approaches have been described and the analysis of segregation of the variant in affected family members has been shown to be very powerful [3]. Co-segregation analysis is usually done with statistical software for linkage analysis in pedigrees [10,11] but a major disadvantage of the currently used methods is that they are not user-friendly and ignore the precise age of onset of the disease since either liability classes or a constant penetrance in carriers is used. We have therefore developed an easy to use method which calculates the likelihood ratio with penetrance as a function of age of onset. The analysis requires at least two genotyped persons with information on gender and present age or age onset for breast and/or ovarian cancer. The software we developed has some similarity with the well-known BRCAPRO software and requires similar input [17]. It is written in MATLAB and the application can be run on our website: <http://www.msbi.nl/cosegregation>

### Essential assumptions

One of the critical steps in our analysis is the selection of the proband. In most cases, this will be the youngest affected family member that was tested positive for the UV. The algorithm analyses the segregation of the UV from the proband and the results can differ according to which person is used as proband. In principle it is possible to allow more complex genotypic ascertainment schemes, but in practice it is essential to know who the proband is.

Another important assumption is that "causal" UV's show the same penetrance as the known deleterious mutations in *BRCA1* and *BRCA2*. In our analysis we did not allow for penetrance functions that depend on the particular mutation although it would be easy to redo the calculations under the assumption of mutation-specific penetrance. Since it would reduce the power of our method, such an analysis would only be feasible in large pedigrees with many genotyped individuals. Thirdly, we assume that the UV is not in linkage disequilibrium with an unknown deleterious mutation. Finally, we computed UV specific LR's by simply multiplying the LR's of the families with the same UV under the assumption of homogeneity of the effect. Unfortunately, we do not have data on enough families with the same UV to analyze the heterogeneity properly.

### The algorithm

The algorithm works by complete enumeration of all genotypic configurations under the assumption that the variant is rare. This method is very fast if the pedigree is not too big. In practice we prune non-informative branches by hand before applying the algorithm. An alternative for manual pruning is to combine our complete enumeration with the peeling algorithm of Elston- Stewart [18]. Such

an algorithm would retain all information, would be able to handle large pedigrees and would speed up the computations. We use the algorithm for breast cancer related to *BRCA1* and *BRCA2* mutations, but it can be easily applied to other cancers, provided the penetrance functions of the genes involved are known.

### Information per pedigree

As we pointed out in the results section, the amount of available information varies between pedigrees due to the size and complexity of the pedigree, and within pedigrees due to the variation in phenotype. As in linkage analysis [19], it pays off to select "extreme" individuals for genotyping in those breast cancer families. In our analysis individuals can have an extreme phenotype (very early onset of disease or disease free at old age) and an extreme genotype (distant relatives that are carrier or non-carriers near the proband). Evidence for causality is obtained if phenotype and genotype are concordant (e.g. carriers with early onset disease and non-carriers healthy family members at old age); evidence for neutrality is obtained if they are discordant (e.g. non-carriers with early disease or healthy old age carriers). As shown in the results section, it is straightforward to compute/simulate the LR's that can be obtained if certain individuals are genotyped. Such a simulation analysis can help in deciding whether to genotype at all, and, if so, who to genotype among those who are willing to be genotyped.

### Clinical findings

We have used the algorithm to calculate the likelihood ratio in favour of causality for 3 UVs in *BRCA1* (i.e. p.M18T, p.S1655F and p.R1699Q) and 5 in *BRCA2* (i.e. p.E462G, p.Y2660D, p.R2784Q, p.R3052W and p.R3052Q). Likelihood ratios varied from 0.097 (*BRCA2*, p.E462G) to 230.69 (*BRCA2*, p.Y2660D). Co-segregation analysis on itself was therefore not powerful enough to classify any of these variants, based on the LR threshold for causality of 1,000:1 ( $LR > 1000$ ) and for neutrality of 100:1 ( $LR < 0.01$ ) [3].

Previous studies have suggested that the *BRCA1* p.S1655F variant and the p.R3052W in *BRCA2* are likely to be deleterious (Additional file 1; Table S1). Our LR's of 6.74 and 12.20 respectively are in line with these findings and support the pathogenic nature of these variants.

The p.E462G (LR 0.097) and p.R3052Q (LR 0.22) variants in *BRCA2* have odds below 1, suggestive for being neutral as has been described by others (Additional file 1; Table S1).

For the p.M18T and p.R1699Q variants in *BRCA1*, LR's of 7.98 and 1.43 respectively were not informative enough to classify them, in line with the inconclusive results obtained by other studies (Additional file 1; Table S1).

These might be variants associated with an intermediate risk and depending on the type of analysis, the variants are classified as likely to be deleterious or neutral. Extensive co-segregation analysis in multiple families with these variants using different penetrance assumptions however might be sufficient to classify these variants in the future.

To date, no other information is available on the pathogenicity of the p.Y2660D and p.R2784Q variants in *BRCA2*. We have analysed co-segregation in 3 families with a p.Y2660D variant and found a LR of 230.69 which is highly suggestive for being a deleterious variant. More families have to be analysed before a definitive conclusion regarding the pathogenicity of this variant can be obtained.

A major problem with the use of co-segregation analysis is that extensive co-segregation analysis is rarely performed in these families, as a result of which only a few families could be used for our analysis. Since the cancer risk associated with these variants is unknown, the information that additional family members carry the variant will not change the screening advice. Therefore, clinical geneticist might be reluctant to perform additional DNA screening in these families, since the knowledge that someone carries the variant might cause anxiety although it is not established that it is associated with elevated cancer risk [20]. The main reason to perform DNA mutation screening in UV families should be however, the contribution to the assessment of the pathogenicity of the variant. We have used families in which either neutral or pathogenic *BRCA1* or *BRCA2* mutations segregate, to determine which family members should be typed to give the most informative results. By careful selection of family members to be typed, i.e. early breast cancer cases distantly related to the proband and old healthy individuals nearby, segregation analysis will be a powerful tool to assess the clinical significance of unclassified variants in *BRCA1* and *BRCA2*.

Although functional data might be useful for classification, to date only functional assays are available for a subset of the UVs, namely those that are located in specific domains in the genes (e.g. BRCT domains in *BRCA1*) or those that affect splicing. Furthermore, the results of the functional assays are difficult to quantify and do not necessarily reflect the influence on cancer risk, since *BRCA1* and *BRCA2* are multifunctional. The determination of the clinical significance of UVs should therefore preferably rely on clinical data, such as co-segregation analysis, which is directly related to disease risk and requires few assumptions. A disadvantage might be that most variants are very rare and that large international collaborations will be required to obtain enough informative families with the same UV to obtain  $LR > 1000$ , the threshold for causality. There is one other drawback to the use of co-seg-

regation analysis; the UV might be in linkage disequilibrium with an unknown deleterious mutation in *BRCA1* or *BRCA2*. To reduced this possibility it is preferred that multiple families are included in the co-segregation analysis.

In the case of most unclassified variants, classification will depend on several data sources (i.e. co-occurrence, family history, tumour information, etc.) as have been described by others [3,4]. In line with the approach of Osorio *et al.* [6], we would like to opt for a model which is based on information that is readily accessible in a clinical setting. Clinical observations, such as co-segregation, family history and tumour information (e.g. LOH and CGH analysis), are directly related to cancer risk and relatively straightforward to quantify. Such a model will considerably increase the clinical utility of *BRCA1* and *BRCA2* genetic testing.

Since unclassified variants are also detected in genes underlying other genetic diseases, e.g. hereditary colon cancer, such a model might be applicable more broadly in a clinical genetic setting.

## Conclusion

Co-segregation analysis on itself is in most cases insufficient to prove pathogenicity of an UV. The presented method, with its user-friendly web application, simplifies the use of co-segregation as one of the key features in a multifactorial approach considerably. Users can upload a file containing the pedigree information on the website to calculate the likelihood ratio. Genotypes of distantly related individuals with extreme phenotypes (i.e. very early onset cancer or old healthy individuals) are most informative and give the strongest likelihood ratios for or against causality.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

LM designed the algorithm, performed the statistical analysis and drafted the manuscript. MPV conceived of the study, obtained financial support, coordinated the data acquisition, analyzed and interpreted data and authored the final version of the manuscript. RO, AO, JCO, AHH, NH, ML, MGA, RBL, CJD, JJG, SV, FBH, TAO, EGG, MJB and JTW were responsible for the selection of the UVs, DNA sequencing and the clinical data collection. QH was responsible for the web application of the algorithm. PD participated in the design of the study, obtained financial support and was involved in revision of the final version of the text. CJA participated in the design of the study, obtained financial support, was responsible for data acquisition, analyzed and interpreted data and authored the final version of the manuscript. HCH contributed to the concept and design of the study, supervised the devel-

opment of the algorithm and helped draft the manuscript. All authors read and approved the final manuscript.

## Additional material

### Additional file 1

**Additional Tables and Figures Mohammadi.** This file contains additional tables 1–3 and additional figures 1–5.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2407-9-211-S1.pdf>]

### Additional file 2

**Instruction method Mohammadi.** This document contains a description on the use of the co-segregation analysis website with detailed instruction on the preparation of the input file that contains the pedigree information.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2407-9-211-S2.pdf>]

## Acknowledgements

This work was supported by the Dutch Cancer Society grants 2001–2471 and 2006–3677. The authors would like to acknowledge Dr. J.J. Houwing-Duistermaat for her effort to improve the user-friendliness of the algorithm and R. Hartog for the development of the graphical interphase.

## References

- Goldgar DE, Easton DF, Byrnes GB, Spurdle AB, Iversen ES, Greenblatt MS: **Genetic evidence and integration of various data sources for classifying uncertain variants into a single model.** *Hum Mutat* 2008, **29**:1265-1272.
- Vink GR, Van Asperen CJ, Devilee P, Breuning MH, Bakker E: **Unclassified variants in disease-causing genes: nonuniformity of genetic testing and counselling, a proposal for guidelines.** *Eur J Hum Genet* 2005, **13**:525-527.
- Goldgar DE, Easton DF, Deffenbaugh AM, Monteiro ANA, Tavtigian SV, Couch FJ: **Integrated evaluation of DNA sequence variants of unknown clinical significance: Application to BRCA1 and BRCA2.** *Am J Hum Genet* 2004, **75**:535-544.
- Chenevix-Trench G, Healey S, Lakhani S, Waring P, Cummings M, Brinkworth R, Deffenbaugh AM, Burbidge LA, Pruss D, Judkins T, Scholl T, Bekessy A, Marsh A, Lovelock P, Wong M, Tesoriero A, Renard H, Southey M, Hopper JL, Yannoukakos K, Brown M, Easton D, Tavtigian SV, Goldgar D, Spurdle AB, kConFab Investigators: **Genetic and histopathologic evaluation of BRCA1 and BRCA2 DNA sequence variants of unknown clinical significance.** *Cancer Res* 2006, **66**:2019-2027.
- Easton DF, Deffenbaugh AM, Pruss D, Frye C, Wenstrup RJ, Allen-Brady K, Tavtigian SV, Monteiro AN, Iversen ES, Couch FJ, Goldgar DE: **A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes.** *Am J Hum Genet* 2007, **81**:873-883.
- Osorio A, Milne RL, Honrado E, Barroso A, Diez O, Salazar R, de la Hoya M, Vega A, Benítez J: **Classification of missense variants of unknown significance in BRCA1 based on clinical and tumor information.** *Hum Mutat* 2007, **28**:477-485.
- Carvalho MA, Marsillac SM, Karchin R, Manoukian S, Grist S, Swaby RF, Urmenyi TP, Rondinelli E, Silva R, Gayol L, Baumbach L, Sutphen R, Pickard-Brzosowicz JL, Nathanson KL, Sali A, Goldgar D, Couch FJ, Radice P, Monteiro AN: **Determination of cancer risk associated with germ line BRCA1 missense variants by functional analysis.** *Cancer Res* 2007, **67**:1494-1501.
- Farrugia DJ, Agarwal MK, Pankratz VS, Deffenbaugh AM, Pruss D, Frye C, Wadum L, Johnson K, Mentlick J, Tavtigian SV, Goldgar DE, Couch FJ: **Functional assays for classification of BRCA2 variants of uncertain significance.** *Cancer Res* 2008, **68**:3523-3531.
- Gómez García EB, Oosterwijk JC, Timmermans M, van Asperen CJ, Hogervorst FB, Hoogerbrugge N, Oldenburg R, Verhoef S, Dommering CJ, Ausems MG, van Os TA, Hout AH van der, Ligtenberg M, Ouweland A van den, Luijt RB van der, Wijnen JT, Gille JJ, Lindsey PJ, Devilee P, Blok MJ, Vreeswijk MP: **A method to assess the clinical significance of unclassified variants in the BRCA1 and BRCA2 genes based on cancer family history.** *Breast Cancer Res* 2009, **11**:R8.
- Petersen GM, Parmigiani G, Thomas D: **Missense mutations in disease genes: a Bayesian approach to evaluate causality.** *Am J Hum Genet* 1998, **62**:1516-1524.
- Thompson D, Easton DF, Goldgar DE: **A full-likelihood method for the evaluation of causality of sequence variants from family data.** *Am J Hum Genet* 2003, **73**:652-655.
- den Dunnen JT, Antonarakis SE: **Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion.** *Hum Mutat* 2000, **15**:7-12.
- Welsh PL, King MC: **BRCA1 and BRCA2 and the genetics of breast and ovarian cancer.** *Hum Mol Genet* 2001, **10**:705-713.
- Statement of the American Society of Clinical Oncology: **genetic testing for cancer susceptibility, Adopted on February 20, 1996.** *J Clin Oncol* 1996, **14**:1730-1736.
- Vreeswijk MP, Kraan JN, Klift HM van der, Vink GR, Cornelisse CJ, Wijnen JT, Bakker E, van Asperen CJ, Devilee P: **Intronic variants in BRCA1 and BRCA2 that affect RNA splicing can be reliably selected by splice-site prediction programs.** *Hum Mutat* 2009, **30**:107-114.
- Jonker MA, Jacobi CE, Hoogendoorn WE, Nagelkerke NJD, de Bock GH, van Houwelingen JC: **Modeling familial clustered breast cancer using published data.** *Cancer Epidemiol Biomarkers Prev* 2003, **12**(12):1479-1485.
- Parmigiani G, Berry DA, Aguilar O: **Determining carrier probabilities for breast cancer- susceptibility genes BRCA1 and BRCA2.** *Am J Hum Genet* 1998, **62**:145-158.
- Elston RC, Stewart J: **A general model for the genetic analysis of pedigree data.** *Hum Hered* 1971, **21**:523-542.
- Lebecq J, Putter H, Houwelingen JC: **Score test for detecting linkage to complex traits in selected samples.** *Genet Epidemiol* 2004, **27**:97-108.
- Vos J, Otten W, van AC, Jansen A, Menko F, Tibben A: **The counsellors' view of an unclassified variant in BRCA1/2: recall, interpretation, and impact on life.** *Psychooncology* 2008, **17**:822-830.

## Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2407/9/211/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

